

## REPRÉSENTATIONS – LANGAGES – COMMUNICATION

Jean-Louis LEBRAVE

*Président*

Viviane Arigne

Dominique Bassano

Éric Beaumatin

Jacques Bres

Christian Cave

Anne Condamines

Bernard Comrie

Francis Corblin

Jocelyne Fernandez-Vest

Nabil Hathout

Jean-Luc Minel

Léa Nash

François Pellegrino

Christian Plantin

François Recanati

Stéphane Robert

Jean-Paul Sansonnet

Bernard Teston

Robert Vion

Richard Zuber

- La section 34 est née de la profonde recomposition des sections du Comité national opérée par le redécoupage de 1991. Elle a fait suite à la section 42 « Sciences du langage », elle-même créée en 1982 pour faire apparaître la linguistique comme une discipline autonome au CNRS. Issue d'une vision prospective qui a été confirmée par l'évolution ultérieure, elle a, autour du noyau disciplinaire constitué par la recherche linguistique, intégré les deux champs alors en émergence des sciences cognitives et des technologies de l'information.

Les effets de ces recompositions successives se sont conjugués avec l'évolution de la discipline elle-même au plan national et international pour produire la situation actuelle, dans laquelle la section 34 se caractérise par une forte interdisciplinarité.

- Le cœur de la section 34 est constitué par les sciences du langage, qui ont pour objectif l'étude du langage humain à travers l'analyse de la diversité des langues, de leurs propriétés universelles et de leurs propriétés particulières. Le langage constitue une capacité spécifique à l'espèce humaine : les humains parlent, et ils sont les seuls à le faire. Cette caractéristique confère à la linguistique une place centrale dans les sciences humaines et sociales.

Les principaux domaines d'étude et les champs traditionnels de la section 34 sont les suivants :

- phonétique, phonologie, morphologie, syntaxe, lexicologie, lexicographie, sémantique, pragmatique ;
- histoire, description et typologie des langues ;
- étude des textes et des discours ;
- psycholinguistique et mécanismes cognitifs ;
- sociolinguistique, ethnolinguistique ;
- traitement automatique des langues et ingénierie linguistique.

Tant du fait de la pluralité des domaines que de la multiplicité des langues étudiées, les sciences du langage se caractérisent par une grande diversité d'approches complémentaires, aussi bien théoriques qu'empiriques. Cette situation, qui n'est pas propre à la communauté des chercheurs français, tient au caractère protéiforme de l'objet langage lui-même.

De par sa composition même, la section 34 est ouverte à toutes les sous-disciplines des sciences du langage, et elle est le point de rencontre d'options théoriques multiples. Cette absence d'unification théorique est caractéristique de l'état actuel du développement des sciences du langage, et la diversité des approches du langage et des langues doit absolument être préservée. Elle est en effet la condition de confrontations fécondes, y compris au travers des tensions que cette diversité peut susciter.

• On soulignera à ce propos un certain nombre de caractéristiques originales dont une analyse de conjoncture doit tenir compte.

Les disciplines n'étant pas structurées de la même manière au CNRS et dans l'université, le regroupement de chercheurs et d'unités de recherche qui correspond à la section 34 n'a pas d'équivalent direct dans les structures universitaires. Il apparaît notamment que l'empan de la section 34 est beaucoup plus large que celui de

la section 07 du CNU (« Sciences du langage : linguistique et phonétique générales »). Du fait de la structuration de l'enseignement supérieur par grandes aires linguistiques, de nombreux enseignants-chercheurs qui, pour leur recrutement et leur évaluation, dépendent de sections du CNU autres que la section 7, relèvent de la section 34 du Comité National pour leur activité de recherche.

Le CNRS a joué historiquement un rôle capital dans la recherche linguistique sur des langues peu ou pas représentées à l'université. D'où un déséquilibre qui perdure entre le poids du CNRS dans la recherche et la formation sur les langues « de terrain », et le rôle relativement marginal de l'enseignement supérieur dans ce domaine.

Au sein même du CNRS, les redécoupages successifs n'ont pas entraîné mécaniquement le rattachement de toute la recherche linguistique à la section 42, puis à la section 34. Un certain nombre de chercheurs ont pendant longtemps préféré mener leurs recherches dans le cadre d'équipes polyvalentes consacrées à des aires culturelles particulières. L'évolution actuelle tend à rendre l'appartenance disciplinaire suffisamment prégnante pour que ces chercheurs ressentent la nécessité de rejoindre la section 34. Mais ce mouvement est loin d'être achevé à ce jour.

## **1 - LA SECTION 34, UNE SECTION INTERDISCIPLINAIRE**

Du fait de la place centrale occupée par le langage dans les activités humaines, la section 34 se caractérise par une très forte interdisciplinarité.

## 1.1 L'INTERDISCIPLINARITÉ INTERNE AUX SCIENCES HUMAINES

On soulignera en premier lieu l'interdisciplinarité avec d'autres disciplines des sciences humaines : sociologie, ethnologie, anthropologie, littérature, philosophie, logique, histoire et épistémologie des disciplines, etc. Alors qu'elle reste présente, comme le montre par exemple la porosité des frontières entre la section 34 et les sections 35, 36, 38 et 40, cette interdisciplinarité interne aux sciences humaines est aujourd'hui traitée en parent pauvre. Les raisons en sont multiples : occultation par d'autres courants interdisciplinaires plus visibles, car en interaction avec des disciplines des sciences plus dures et avec d'autres départements du CNRS ; effet de mode inversé – cette interdisciplinarité était privilégiée au temps où la linguistique structuraliste était la discipline phare des sciences humaines ; vieillissement des chercheurs, insuffisamment renouvelés. Il s'y ajoute sans doute aussi l'effet d'une fascination, pas toujours justifiée, pour les disciplines dites dures, supposées être plus structurantes que les sciences humaines et sociales.

Quelle que soit l'importance – incontestable – des autres chantiers interdisciplinaires auxquels les sciences du langage participent, il convient de ne pas oublier ces chantiers internes, dont l'importance tient à la place du langage comme activité sociale et comme pratique culturelle. C'est l'un des messages qu'on souhaite faire passer avec ce rapport. Il serait en particulier très dommageable que la raréfaction des postes et l'intensification des affichages condamnent de fait cette interdisciplinarité à une mort lente ou soit abandonnée aux recherches universitaires.

## 1.2 L'INTERDISCIPLINARITÉ GLOBALE

Plus visible, et faisant l'objet d'une grande attention de la part des instances de direction du CNRS, est l'interdisciplinarité que la section 34

a construite depuis sa création en 1991 avec des disciplines relevant d'autres départements du CNRS. Les interactions concernent ici d'une part la psycholinguistique et les sciences cognitives, d'autre part le traitement automatique des langues et l'ingénierie linguistique, développées en relation étroite avec des chercheurs et des laboratoires appartenant à d'autres sections, particulièrement les sections 29 et 07.

- Au CNRS, la psycholinguistique et la neuropsycholinguistique relèvent, on le sait, de la section 29 et des Sciences de la Vie, alors qu'elle est intégrée aux sciences humaines et sociales dans les découpages universitaires. Cette séparation somme toute récente a eu des effets de clivage au sein des communautés concernées, qui ont pu engendrer des formes d'incompréhension (insuffisante formation à la mise en place de protocoles expérimentaux, reprochée aux linguistes par les psychologues ; faiblesse théorique des modèles linguistiques sous-jacents aux procédures d'expérience, reprochée aux psychologues par les linguistes).

Heureusement, grâce à la politique de recrutements croisés mise en place depuis quelques années par le CNRS, l'évolution récente tend à résorber progressivement ces clivages en rétablissant les conditions d'un dialogue fructueux entre les deux communautés.

D'une manière beaucoup plus récente, et encore quasi expérimentale, on assiste au développement progressif de coopérations entre des chercheurs relevant de la section 34 et les neurosciences intégratives, notamment avec la présence de plus en plus affirmée de l'imagerie cérébrale hors du champ strict des neurosciences. On citera en particulier les chantiers ouverts par les chercheurs du GDR animé par Stéphane Robert.

Plus généralement, la section 34 est très impliquée dans le champ des sciences cognitives. Fortement soutenu par des mesures incitatives depuis une vingtaine d'années, ce domaine de recherche continue à être en pleine expansion, comme le montrent à la fois l'engagement d'un nombre croissant de chercheurs et la poursuite de la politique volontariste des

tutelles (création en 2003 de la commission interdisciplinaire 45). Dans le cadre de la section 34, deux courants principaux sont impliqués dans ces recherches. Schématiquement, le premier rassemble des chercheurs qui se réclament de la philosophie de l'esprit anglo-saxonne, des grammaires formelles d'inspiration chomskyenne et d'hypothèses innéistes sur les propriétés universelles des langues. Le second adopte le point de vue des grammaires cognitives et s'enracine davantage dans la tradition de la recherche linguistique européenne. Il convient de préserver dans les années à venir cette diversité, qui suscite des discussions, voire des controverses extrêmement fécondes.

On notera enfin à ce propos que la communauté française des chercheurs en intelligence artificielle et, plus généralement, en informatique, qui était très présente dans l'émergence des sciences cognitives en France à la fin des années 1980, paraît occuper actuellement une position plus en retrait.

- Les sciences et technologies de l'information constituent l'autre grand domaine d'interface de la section 34. Les coopérations se sont ici mises en place très tôt – comme le montre par exemple le rôle qu'a joué pour la recherche linguistique le modèle de la communication élaboré par les ingénieurs du téléphone au début des années 50. Elles ont largement bénéficié du développement de domaines très porteurs, qu'il s'agisse du traitement de la parole, des interactions homme-machine, ou de la traduction automatique, relayée ensuite par le traitement automatique des langues et la linguistique informatique.

Parmi les faits marquants des dernières années, on citera en particulier le remarquable essor de la linguistique de corpus. La restructuration de l'INALF autour d'une meilleure articulation avec les équipes de recherche STIC nancéennes et la création de la fédération de linguistique française sont ici plus que des symboles.

On trouvera en annexe des contributions illustrant quelques points de vue sur ces recherches au croisement de la section 34 et des STIC.

- À juste titre, cette interdisciplinarité avec les sciences du vivant et les STIC a été fortement mise en avant depuis le redécoupage du Comité National de 1991. Parmi les sciences humaines et sociales, la linguistique occupe à ce titre une position exemplaire.

## **2 – POUR UN DÉVELOPPEMENT ÉQUILIBRÉ**

Si l'interdisciplinarité est un atout majeur pour la section 34, il convient toutefois qu'elle ne se développe pas en étouffant le cœur de la discipline. Du fait de la politique volontariste pratiquée par le CNRS depuis une quinzaine d'années – et notamment des fléchages systématiques de postes « interdépartements » –, une part importante des recrutements récents ont été effectués au bénéfice des interfaces. Cette expansion très positive doit toutefois rester contrôlée et ne pas s'effectuer au détriment d'autres domaines des sciences du langage.

On citera deux exemples qui, à des titres différents, méritent l'attention : la « linguistique des langues », et le vaste domaine de la textualité, du discours et des pratiques langagières.

### **2.1 LA LINGUISTIQUE DES LANGUES**

Elle a connu une profonde mutation depuis une quinzaine d'années, qui s'est traduite par une profonde restructuration des unités de recherche du domaine. Schématiquement, on peut résumer cette évolution en posant qu'on est passé d'une linguistique de terrain issue de la période coloniale à une linguistique beaucoup plus centrée d'une part sur les problèmes de typologie et la recherche d'universaux, et d'autre part sur les problèmes cognitifs posés par la diversité des langues.

Après une période de relatif repli national, cette communauté est de plus en plus ouverte vers l'extérieur, et elle occupe une place de choix dans la recherche internationale. En témoigne – mais ce ne sont que des exemples parmi d'autres – la réémergence récente du problème de l'origine du langage et des langues, qui constitue actuellement un domaine très interdisciplinaire en pleine expansion, ou la prise de conscience de plus en plus aiguë de la disparition d'un nombre croissant de langues et des problèmes que pose la sauvegarde de cette partie du patrimoine de l'humanité.

On signalera toutefois un élément préoccupant, lié au vieillissement des chercheurs recrutés dans les années 1970. Les vagues de départs en retraite en cours priveront ce champ de chercheurs de qualité, souvent seuls spécialistes de leur domaine, sans que les flux de recrutement actuels permettent leur remplacement dans des conditions satisfaisantes. Le problème est aggravé par des facteurs presque mécaniques : la durée de la formation étant plus longue dans ce domaine – et le CNRS étant souvent le seul acteur pour des langues peu ou pas enseignées à l'université –, il est souvent difficile, voire impossible pour les jeunes chercheurs de satisfaire aux critères d'âge imposés pour le recrutement comme CR2. Il convient d'élaborer rapidement une politique de formation et de recrutement susceptible de corriger les effets induits par cette disparité, qui sinon risquent à terme de transformer la linguistique des langues en secteur sinistré.

## ANNEXE

### LE TRAITEMENT AUTOMATIQUE DES LANGUES

Le traitement automatique de la langue est une thématique qui a son propre objectif tout en entretenant des liens étroits avec

## 2.2 TEXTUALITÉ, DISCOURS, PRATIQUES LANGAGIÈRES

Au même titre que les enjeux cognitifs, la description des langues et la typologie, ce domaine était l'un des axes prioritaires du département dans les orientations stratégiques publiées en 2000. Ce texte n'a rien perdu de son actualité : « Les sciences humaines et sociales attendent de la linguistique qu'elle développe des méthodes d'analyse permettant de comprendre ce qui se joue au niveau [des productions langagières]. Les productions monogales et les interactions verbales doivent faire l'objet d'une approche théorique et méthodologique unifiée [...] Les productions écrites et orales doivent être appréhendées en respectant leurs variabilités et leurs relatives spécificités ».

Il est souhaitable que les regroupements d'unités en cours contribuent à mettre en œuvre ce programme en permettant aux recherches sur les textes, les discours et les pratiques langagières d'atteindre une masse critique.

L'étude de la production verbale, orale et écrite, longtemps négligée par la recherche linguistique, paraît susceptible de connaître des développements prometteurs avec la création d'un GDR interdisciplinaire réunissant des linguistes et des psycholinguistes autour des problèmes de la production écrite.

différentes disciplines des sciences humaines et tout particulièrement la linguistique. Il conviendra de mieux étudier les interactions possibles entre TAL et linguistique, dans une perspective qui peut se décliner de la manière suivante :

Pour le TAL :

– rôle de l'analyse de données réelles : corpus généraux ou spécialisés, dialogues, etc.,

- définition de modèles rendant compte de ces données,

- conception d'outils prenant en compte des connaissances linguistiques de plus en plus fines,

- couplage d'approches probabilistes et linguistiques.

Pour la linguistique :

- rôle de l'analyse de corpus dans les études linguistiques : quelle possibilité de théorisation ?,

- rôle des outils de TAL dans l'analyse de corpus : possibilités et limites,

- production de ressources pour améliorer les performances des outils de TAL,

- formation des linguistes aux outils et à leurs principes d'analyse.

On retrouve tout ou partie de ces perspectives dans les différents domaines du traitement automatique de la langue.

nant. Il est en effet indispensable de prendre en compte simultanément plusieurs domaines de l'analyse linguistique (phonétique, prosodie, syntaxe, sémantique, pragmatique) en expliquant leurs interactions, de façon cohérente et systématique. Le TAL peut participer à ce mouvement tout d'abord en contribuant à la mise au point de modèles théoriques : les approches basées sur les contraintes permettent par exemple de fournir un cadre formel et computationnel à ces modèles interactionnistes. De plus, les enjeux applicatifs sont importants : traitement de la langue parlée (reconnaissance et synthèse de la parole, dialogue oral, etc.), traitement de textes tout-venant (analyse de corpus), traitement de l'information, de son organisation et sa transmission.

La mise au point de systèmes s'appuyant sur ce type d'approche intégrée constituerait alors une véritable plate-forme d'observation et de description des processus cognitifs mis en œuvre dans la langue, à la fois du point de vue de la production et de la perception en permettant en particulier d'examiner le rôle de chacun des domaines par rapport aux autres.

## **1 MODÈLES THÉORIQUES POUR LES LANGUES**

Le TAL a pendant de nombreuses années été traversé par deux courants présentés comme antagonistes : l'un relevant de l'ingénierie, et l'autre s'appuyant sur des modèles théoriques. Dans ce dernier cas, le TAL était plus utilisé comme validation de la cohérence d'une théorie que comme véritable outil d'investigation. Ces approches sont désormais complémentaires : les théories linguistiques tirent parti des résultats de l'analyse de corpus, de même que les applications (y compris les approches probabilistes) s'appuient sur des données linguistiques organisées.

Un des enjeux majeurs du TAL réside aujourd'hui dans le traitement de matériel linguistique non normé : langue parlée, textes tout-venant, etc. Les modèles théoriques joueront dans cette perspective un rôle détermi-

## **2 INTERACTIONS LANGAGIÈRES**

Le dialogue est pour les humains le lieu de l'apprentissage des connaissances et des langues. C'est un processus social dans lequel les machines sont maintenant partie prenante. Les interactions langagières entre les humains et les machines et entre les humains à travers les machines font l'objet de nombreux travaux récents qui visent à prendre en compte la multimodalité (parole, texte, gestes et mimiques), et la cognition située et distribuée. La communication entre agents d'un SMA (Systèmes MultiAgents : agents logiciels, robots, humains) commence à être vue comme le paradigme englobant pour l'étude des interactions langagières. Les avancées dans ce domaine amènent à prendre en compte de nombreux modèles comme les modèles de la langue en linguistique, les modèles de

l'interaction et de l'acquisition du langage en psycholinguistique, les modèles du fonctionnement cognitif et langagier en sciences cognitives, les modèles des protocoles sociaux qui structurent les conversations. Construire des modèles computationnels qui puissent en rendre compte, mais aussi les évaluer et les faire évoluer est un objectif important des prochaines années. Les objectifs socio-économiques attendus sont importants : instrumenter la communication humaine (mémoire, gestion du temps), faciliter les interactions humains-machines (expression, ergonomie, esthétique), définir le statut social des machines dans l'interaction (substituts, mandataires, intermédiaires), les droits et devoirs qui s'y attachent en tant que « parlant ».

L'interaction pluridisciplinaire est essentielle pour mener à bien ses projets, tant pour l'analyse de corpus dans des conditions nouvelles (mail, chats, forums, mondes virtuels) que pour l'évolution des artefacts langagiers. Elle permet de croiser l'analyse des corpus et des résultats et de construire des modèles reconnus tant par les informaticiens que par les chercheurs en sciences humaines. Les objectifs scientifiques actuels sont de définir des modèles des langues communs pour l'analyse et la génération, des modèles de l'activité langagière qui permettent l'articulation entre les modalités d'expression, des modèles de l'articulation entre activité langagière et non langagière dans des tâches finalisées. Ces modèles devront être testés, comparés, et validés de manière interne et externe dans une perspective interdisciplinaire.

### **3 ACQUISITION ET REPRÉSENTATION DE CONNAISSANCES SÉMANTIQUES**

L'acquisition de connaissances sémantiques et leur représentation constituent un des enjeux majeurs du TAL, en lien avec l'Ingénierie des Connaissances. Elles peuvent être traitées avec différents points de vue : par domaine de connaissances (terminologie) vs pour une langue

(lexicologie) et avec un point de vue plus ou moins formel. L'acquisition de ces connaissances peut être réalisée sur des bases introspectives ou bien auprès d'experts ou encore à partir de textes. Chaque étape d'élaboration de ces données (acquisition, modélisation, formalisation) peut bénéficier de connaissances linguistiques (lexicologie, terminologie, dialogue, sémantique de corpus, analyse de discours, etc.). Mais une réflexion globale doit se développer sur les méthodes, les compatibilités entre approches linguistiques et informatiques, sur des questions aussi diverses que l'hétérogénéité des données, la mise au jour et le traitement de la polysémie, la compatibilité entre les modèles de données, le développement et l'utilisation d'outils, la prise en compte d'applications réelles et leur rôle sur l'interprétation sémantique. Dans le cas de l'acquisition à partir de textes, le lien avec l'analyse de corpus et la Recherche d'Information est patent et doit être mieux pris en considération. Ce rapprochement avec les besoins en ingénierie des connaissances peut permettre une réelle valorisation des connaissances et des compétences en linguistique.

## **4 ACCÈS À L'INFORMATION TEXTUELLE**

Les dépôts d'informations textuelles sont en croissance exponentielle sur le Web comme dans les intranets d'entreprise. Leurs documents sont hétérogènes (formats et genres multiples) et souvent multimédia ; et la mondialisation s'accompagne du multilinguisme.

Pour faire face à ce défi, la Recherche d'Information (traitement de masse, essentiellement statistique, faible profondeur) et le TAL (analyse du contenu, volumes plus limités) sont voués à la convergence. Le TAL aide à regrouper selon le sens (morphologie, synonymes et hyperonymes). Il permettra de distinguer (désambiguïsation syntaxique et sémantique) et de sortir du « sac de mots » pour prendre en compte la textualité (dépendances syntaxiques, actants, résolution d'anaphores).

L'Extraction d'Information a ouvert la voie pour aspirer des informations ciblées de documents de domaine et genre fixés. Les systèmes de Question-Réponse (question en langage clair et domaine ouvert, réponses sous forme de courts extraits de textes), qui permettront à l'utilisateur d'aller plus rapidement au but dans une recherche, constituent le fer de lance du retour d'une dimension linguistique dans l'accès à l'information : analyse de questions naturelles, analyse robuste de textes, résolution d'anaphores, appariement sémantique et inférence.

Citons encore la Recherche d'Information Trans-Langue (requête monolingue, documents multilingues), qui ouvre le champ du multilinguisme à la fraction grandissante de la population qui connaît, même passivement, une seconde langue. L'ancêtre inspirateur (Traduction Automatique), le développement d'auxiliaires techniques (alignement de corpus parallèles ou comparables) participent du même élan.

Se fait toujours cruellement sentir le besoin de ressources sur le français (grands lexiques, bases de connaissances morphologiques, de synonymes et d'hyperonymes, corpus annotés) et d'outils de base (analyseurs robustes, reconnaissance d'entités nommées, etc.) qui devraient être librement disponibles pour que la recherche capitalise et se concentre sur l'innovation.

Enfin, l'enjeu du TAL dans l'accès à l'information textuelle est de joindre profondeur d'analyse avec volume et diversité des documents traités. Il est parallèle à la nécessaire alliance entre méthodes symboliques (modélisation linguistique) et numériques (statistiques).

L'appropriation des méthodes statistiques et d'apprentissage doit être favorisée. Le contact méthodologique avec le traitement de la parole, où cette alliance existe avec une dominante numérique, sera mutuellement bénéfique. Simultanément, la linguistique a un rôle à jouer pour guider la conception de modèles plus élaborés, en particulier dans la prise en compte de la textualité (genres textuels).

## 5 CORPUS ET OUTILS LINGUISTIQUES

De nombreux travaux en linguistique mettent en œuvre des corpus selon deux approches souvent complémentaires : l'une qui consiste à vérifier dans des textes pris « en vrac » des hypothèses de fonctionnement, l'autre qui consiste à utiliser un corpus comme référence pour étudier certains phénomènes linguistiques en lien avec un genre textuel. Pour faciliter ce type d'analyse, la palette des outils et ressources accessibles pour le français est devenue relativement importante : étiqueteurs morpho-syntaxiques, lexiques généralistes ou restreints, corpus de diverses natures : littéraires, presses, spécialisés. Les données annotées plus finement (alignement, anaphores, etc.) sont venues s'ajouter plus récemment. Ces ressources, encore insuffisantes sont par ailleurs peu utilisées car mal recensées et peu accessibles, souvent pour des raisons financières. Par ailleurs, la nécessité de connaissances en TAL restreint leur utilisation par les linguistes. Aussi, l'intégration de ces ressources dans la formation des jeunes chercheurs constitue un objectif crucial. Formation technique d'abord : connaissance des performances des outils, de la fiabilité de l'annotation tant humaine qu'automatique, maîtrise des standards d'annotation, pratique de l'adaptation d'un jeu de données à un objectif de recherche déterminé, etc. Formation aux méthodes de constitution et d'utilisation des corpus ensuite : utilisation et adaptation, à partir d'une hypothèse, des ressources disponibles ; mise en œuvre d'une approche quantitative adéquate.

Deux éléments convergent en faveur d'un développement des études linguistiques à partir de corpus : la demande sociétale et l'évolution théorique d'une discipline qui s'intéresse de plus en plus aux données attestées. Cette mutation constitue ainsi un des défis majeurs de la linguistique et elle doit être accompagnée de toutes les façons possibles : interrogation épistémologique sur le rôle des corpus, développement de méthodes, formation des chercheurs, mise à disposition de ressources, etc.



## **6 INTERACTIONS ENTRE LA SECTION 34 ET LA SECTION 07 (F. PELLEGRINO)**

Dans le précédent rapport de conjoncture (1996), le traitement automatique des langues et les sciences cognitives étaient mis en avant comme étant les domaines où l'interaction interdisciplinaire entre les sections 07 et 34 était particulièrement active et prometteuse. Quelques 6 ans après, ce constat est plus que jamais d'actualité.

À l'heure actuelle et plus encore que par le passé, on assiste à des échanges productifs entre les communautés du traitement du signal et des images, de l'informatique et de la linguistique visant aussi bien à la conception de systèmes automatiques qu'à l'avancée de la réflexion théorique, non seulement en linguistique mais aussi dans le cadre plus vaste des sciences de la cognition.

### ***Traitement automatique de la parole***

La conception de systèmes de dialogue homme-machine performants implique une collaboration efficace entre les sections 07 et 34, tant en reconnaissance de parole, où la levée des ambiguïtés et le défi de la multilingualité imposent de plus en plus de prendre en compte des connaissances linguistiques qu'en synthèse automatique, où, au-delà de l'intelligibilité, l'amélioration de la qualité de l'énoncé bénéficie de l'ajout de variabilité artificielle et de la prise en compte de phénomènes prosodiques locaux et globaux.

Ces thèmes de collaboration phares ne doivent pas faire oublier d'autres applications pluridisciplinaires du traitement de la parole comme l'identification automatique des langues ou l'indexation de documents sonores et/ou audiovisuels à partir du contenu, enjeux majeurs de la société de l'information multilingue du XXI<sup>e</sup> siècle.

Ajoutons également que les progrès réalisés en traitement du signal permettent d'initier des collaborations prometteuses dans le domaine des aides auditives, tant pour l'amélioration de l'intelligibilité de la parole par manipulation du signal que pour le développement d'implants cochléaires « intelligents ».

### ***Linguistique***

La confrontation des approches développées dans les sections 07 et 34, guidées à la fois par la volonté d'obtenir des systèmes automatiques performants et d'atteindre une meilleure description et compréhension des processus linguistiques, amène un éclairage nouveau sur des questions traditionnelles. On peut citer par exemple le renouveau des approches visant à la compréhension des mécanismes de la prosodie et du rythme ou encore la réflexion en cours sur la caractérisation de la variabilité (acoustique, phonétique, etc.) du signal de parole et sur son rôle dans la communication parlée et dans l'évolution du langage et des langues.

Sur le thème connexe de l'origine du langage, la dernière décennie a vu également émerger des collaborations fructueuses entre informaticiens et linguistes basées sur l'application de paradigmes d'auto-organisation et de systèmes multi-agents.

Ce dernier point, ainsi que les recherches menées sur les interactions communicationnelles entre agents humains et/ou artificiels dépassent les enjeux stricts de la linguistique et entrent pleinement dans le vaste cadre pluridisciplinaire de la communication langagière et des sciences de la cognition.

### ***Sciences de la cognition***

La compréhension des mécanismes de traitement du langage par l'être humain est un enjeu majeur tant sur le plan scientifique que pour ses retombées technologiques et sociales (traitement des troubles du langage, pour ne citer qu'un point). La complexité

de ce domaine particulièrement vaste nécessite des collaborations intenses entre les sections 07, 29 et 34, de manière à confronter des modèles automatiques et théoriques aux résultats expérimentaux.

Outre les techniques d'acquisition et de traitement d'images cérébrales fonctionnelles qui révolutionnent actuellement l'investigation scientifique, la section 07 apporte également

son expertise dans le domaine de la modélisation aussi bien en neurosciences qu'en sciences cognitives : organisation et modélisation des mémoires langagières, simulation de lésions cérébrales sur des réseaux de neurones artificiels, intégration de l'aspect temporel dans les modèles de mécanismes cognitifs, etc.